Exponential inequalities for counting processes

Patricia REYNAUD-BOURET
Georgia Institute of Technology
E-mail: Patricia.Reynaud-Bouret@ens.fr

1st November 2002

Abstract

We derive for counting processes some equivalents of classical exponential inequalities existing in the n-sample framework. The supremum of centered integrals is our principal interest: we prove an inequality which resembles Talagrand's inequality for empirical processes. This result is especially useful when we want to do some model selection for estimating the intensity of the process.

AMS Classification: 60E15, 60G55.
Keywords: Exponential inequalities, point or counting process, supremum.

1 Introduction

Let $(N_t)_{t\geq 0}$ be a counting process i.e. a random increasing piecewise constant function with $N_0 = 0$ and with jumps equal to 1. We want to derive for these processes some exponential inequalities which resemble inequalities already existing in the n-sample framework. Actually the counting processes are really useful to model a huge amount of situations which have biomedical origins (see [1]); they can also model breakdowns, earthquakes... To manage such kind of data, some powerful probabilistic tools such as the exponential inequalities are necessary. This is the aim of this paper.

Let $(\mathcal{F}_t)_{t\geq 0}$ be the filtration generated by $(N_t)_{t\geq 0}$ and $(\Lambda_t)_{t\geq 0}$ be the compensator of the counting process N, i.e. the nondecreasing function such that $(M_t = N_t - \Lambda_t)_{t\geq 0}$ is a martingale (for precise definitions, see for instance [4]).

A known result for martingales is the following one [6, Theorem 23.17].

Theorem 1. *Let $(Z_t)_{t\geq 0}$ be a local martingale with $Z_0 = 0$ and with jumps bounded by b smaller than 1. Suppose that a.s. $< Z >$ is bounded by 1. Then there exists some constant C such that for all positive r,*

$$\mathbb{P}(\sup_{t\geq 0} Z_t \geq r) \leq C \exp\left(-\frac{1}{2} r \log(1 + rb)/b\right).$$

This result exists also for a supremum on $[0, T]$ for fixed positive T by stopping the martingale. We can apply this result to the particular martingale $(Z_t)_{t\geq 0}$ defined by:

$$\forall t \geq 0, Z_t = \int_0^t H_s dM_s = \int_0^t H_s(dN_s - d\Lambda_s), \tag{1.1}$$

where $(H_t)_{t\geq 0}$ is a locally bounded predictable process with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$. Let T be some positive fixed time eventually infinite. If we assume that there exist b and v positive deterministic constants such that

- $(H_t)_{T\geq t\geq 0}$ is bounded by b,
- $\int_0^T H_t^2 d\Lambda_t$ is bounded by v,

then, applying Theorem 1, we can easily prove that for all positive x:

$$\mathbb{P}\left(\sup_{[0,T]} Z_t \geq x\right) \leq C\exp\left(-\frac{1}{2}\frac{x}{b}\log\left(1+\frac{xb}{v}\right)\right). \tag{1.2}$$

We can compare this inequality with those already existing in the n-sample framework. For this, we need some correspondence between the n-sample framework and the counting processes framework. Let $X_1, ..., X_n$ be n i.i.d. real random variables with law $d\mathbb{P}$.

- dN_t is a random measure which corresponds to the empirical measure
$$nd\mathbb{P}_n = \delta_{X_1} + ... + \delta_{X_n}. \tag{1.3}$$
- $d\Lambda_t$ corresponds to the expectation measure $nd\mathbb{P}$.
- $(H_t)_{t\geq 0}$ corresponds to some deterministic function f.
- $\int_0^T H_t^2 d\Lambda_t$ corresponds to $\int f^2(x)nd\mathbb{P}(x)$ which has the same order of magnitude as
$$n\mathrm{Var}(f(X_1)) = \mathrm{Var}(\int f(x)nd\mathbb{P}_n(x)).$$

With this equivalence, we see that Inequality (1.2) with $C=1$ is very close to the well known Bennett's inequality available for a sum of independent bounded variables. The only difference is that here there is a supremum in time, which is actually just a refinement: with martingales, we can use stopping times to easily obtain the inequalities for supremum in time from the inequalities without supremum in time.

Moreover we can prove Inequality (1.2) with $C=1$ in the particular case of (1.1) by applying the same kind of computations as in Bennett's inequality. This is done very quickly in the next Section under the forthcoming assumption:

Assumption 1. *The compensator $(\Lambda_t)_{t\geq 0}$ is absolutely continuous and a.s. finite on $[0,T]$.*

The first part means that $d\Lambda_s$ is absolutely continuous with respect to the Lebesgue measure ds and the second part implies that N has a.s. a finite number of jumps. The processes with at most n jumps (n fixed) but also time Poisson processes with finite mean measure on $[0,T]$ verify this assumption. It means that there is no accumulation point for the jumps on $[0,T]$.

The other classical exponential inequality in the i.i.d. framework is Bernstein's one. As $\int_0^T H_s^k d\Lambda_s$ corresponds to $\sum_{i=1}^n E(X_i^k)$, there exist a similar result for the martingale given by (1.1), due to S. van de Geer [14] in a more general framework:

Proposition 1. *Let $(Z_t)_{t\geq 0}$ be the process:*

$$\forall t\geq 0, \quad Z_t = \int_0^t H_s dM_s,$$

where $(H_t)_{t\geq 0}$ is a predictable process. Let T be a positive real number. If there exists c and v deterministic positive, such that

$$\forall k \geq 2, \left| \int_0^T H_s^k d\Lambda_s \right| \leq c^{k-2} v \frac{k!}{2},$$

under Assumption 1, then for all positive u,

$$\mathbb{P}\left(\sup_{[0,T]} Z_t \geq \sqrt{2vu} + cu \right) \leq \exp(-u).$$

As Inequality (1.2), Proposition 1 is proved very quickly in the next Section because one gets again a proof which is simpler and very similar to Bernstein's inequality when we assume that Z is given by (1.1) under Assumption 1. These proofs are based on the existence of an exponential super-martingale.

Let us give an easy but useful corollary.

Corollary 1. *Let $(Z_t)_{t\geq 0}$ be the process defined by:*

$$\forall t \geq 0, \quad Z_t = \int_0^t H_s dM_s,$$

where H is a predictable process. Let T be a positive real number. If there exist b, v deterministic positive constants such that, on $[0,T]$ $(H_t)_{t\geq 0}$ and $(< Z >_t)_{t\geq 0}$ are bounded respectively by b and v then under Assumption 1, for all positive u,

$$\mathbb{P}\left(\sup_{[0,T]} Z_t \geq \sqrt{2vu} + \frac{b}{3}u \right) \leq \exp(-u).$$

This is a corollary of Inequality (1.2) with $C = 1$ and also a corollary of Proposition 1. This result corresponds in the i.i.d. framework to this inequality which is a consequence of Bernstein's inequality:

Proposition 2. *Let $X_1, ..., X_n$ be n i.i.d. random variables bounded by b with mean 0 and let $S_n = \sum_{i=1}^n X_i$. Then for all positive u,*

$$\mathbb{P}\left(S_n \geq \sqrt{2\mathrm{Var}(S_n)u} + \frac{b}{3}u \right) \leq \exp(-u).$$

An other classical result in the i.i.d. framework is Talagrand's inequality. This is a generalization of the previous result for a suprema of empirical processes. Its simplest form is E. Rio's one [11]:

Theorem 2 (Talagrand's inequality). *Let $X_1, ..., X_n$ be i.i.d. random variables with values in $(\mathbb{X}, \mathcal{X})$. Let $\{\psi_a, a \in \mathcal{A}\}$ be a countable family of real measurable functions with values in $[-b; b]$. Let*

$$Z = \sup_{a\in\mathcal{A}} \sum_{i=1}^n [\psi_a(X_i) - \mathbb{E}(\psi_a(X))] \ \text{ and } v = n \sup_{a\in\mathcal{A}} \mathrm{Var}(\psi_a(X_1)).$$

Then for all positive u

$$\mathbb{P}\left(Z \geq \mathbb{E}(Z) + \sqrt{(2v + 4b\mathbb{E}(Z))u} + (b/2)u \right) \leq \exp(-u).$$

This inequality implies that for all ε and u positive,

$$\mathbb{P}\left(Z \geq (1+\varepsilon)\mathbb{E}(Z) + \sqrt{2vu} + ((1/2)+\varepsilon^{-1})bu\right) \leq \exp(-u). \tag{1.4}$$

When we compare it to Proposition 2, we see that v plays the role of the variance term because of its position in the quadratic term of the deviation. Up to the constants, this is consequently really the generalization of Proposition 2.

It is possible to obtain a corresponding result for counting processes. A result is already available for the Poisson processes (i.e. when Λ is deterministic) [9, Corollary 1]. But we prove in Section 3 a result available for more general counting processes.

Proposition 3. *Let $\{(H_{a,t})_{t\geq 0}, a \in \mathcal{A}\}$ be a countable family of predictable processes. Let*

$$\forall t \geq 0, Z_t = \sup_{a\in\mathcal{A}}\left[\int_0^t H_{a,s}dM_s\right].$$

Let T be a positive real number. Under Assumption 1, the process $(Z_{t\wedge T})_{t\geq 0}$ has a positive nondecreasing compensator $(A_t)_{t\geq 0}$ and one has:

(a) if the $|H_a|$ and $\int_0^T \sup_{a\in\mathcal{A}}[H_{a,s}^2]d\Lambda_s$ are bounded respectively by b on $[0,T]$ and v, both deterministic positive then for all u positive,

$$\mathbb{P}\left(\sup_{[0,T]}(Z_t - A_t) \geq \sqrt{2vu} + \frac{1}{3}bu\right) \leq \exp(-u),$$

(b) if there exists c and v deterministic positive, such that

$$\forall k \geq 2, \left(\int_0^T \sup_{a\in\mathcal{A}}|H_{a,s}|^k d\Lambda_s\right) \leq c^{k-2}v\frac{k!}{2}$$

then for all u positive,

$$\mathbb{P}\left(\sup_{[0,T]}(Z_t - A_t) \geq \sqrt{2vu} + cu\right) \leq \exp(-u).$$

Let us notice at first that here we manage a quantity which is a supremum in time but also in the other index a. Therefore, this is not an obvious consequence of Corollary 1 using stopping times and really requires other techniques.

Secondly, with the correspondence given in (1.3), it seems at first look that this new inequality is in some sense stronger than Theorem 2 in the i.i.d. framework or Corollary 1 of [9] in the Poisson framework (which are up to constants equivalent if we use the correspondence given in (1.3)): we can manage random (predictable) functions and one has also a "moment" version (see (b)), which doesn't suppose a bound on the family of functions to integrate (or to sum in the i.i.d. framework).

The presence of the $\sup_{[0,T]}$ is just a refinement due to the martingale structure but does not affect the orders of magnitude.

However we loose some important fact in these orders with respect to Talagrand's inequality: let us compare the variance term v for each case. In Theorem 2, v can be seen as

$$v = \sup_{a\in\mathcal{A}}\mathbb{E}(\sum_{i=1}^n(\psi_a(X_i) - \mathbb{E}(\psi_a(X_i))^2).$$

The supremum is outside the sum but in Proposition 3(a), it lies inside the integral and is consequently of bigger order.

This phenomenon was already underlined by P.-M. Samson [13]. He recovers Talagrand's inequality for Φ-mixing up to this exchange between the supremum and the sum. For the Poisson processes, L. Wu [15] and C. Houdré and N. Privault [5] used martingales approach to derive exponential inequalities for very general functionals of the process. When we apply these inequalities to supremum, this exchange also appears in the variance term. To have supremum on the left hand side in the Poisson case [9], we need some techniques using the infinitely divisible property of the Poisson process. Consequently it seems that the exchange between supremum and sum (or integral) can be made only when there exists some independence property in the problem. In a general setup like counting processes, we have not such type of results.

Remark: All the results (Inequality (1.2), Proposition 1, Corollary 1 and Proposition 3) are also true if the assumptions on c, v and b are not satisfied on the whole probability space but on an event $\mathcal{C}$: the bound remains the same for the probability of the intersection of $\mathcal{C}$ and "$\sup Z \geq ...$". This is a refinement proved by stopping the martingales, as $\mathcal{C}$ can be described in term of stopping times.

The last part is devoted to explanations about the statistical interest of such suprema and an useful application of these inequalities to the χ^2-type statistics which appear naturally in model selection. We give also some orders of magnitude in a simple case in order to better explain the difference between the variance terms.

2 Basic exponential inequalities

We start with the existence of some super-martingale.

Proposition 4.
Let $(H_t)_{t\geq 0}$ be a locally bounded predictable process and $(Z_t)_{t\geq 0}$ be defined for all positive t by $Z_t = \int_0^t H_s dM_s$. Let $\phi(u) = e^u - u - 1$, for all u and finally let

$$\forall t \geq 0, \quad E_t = \exp\left(\lambda Z_t - \int_0^t \phi(\lambda H_s) d\Lambda_s\right).$$

Let T be a positive real number. Let I be an interval such that for all λ in I, $\int_0^T e^{\lambda H_s} d\Lambda_s$ is a.s. finite, then under Assumption 1 $(E_{t\wedge T})_{t\geq 0}$ is a super-martingale and for all τ less than the stopping time T, $\mathbb{E}(E_\tau)$is less than 1.

Proof. Let us fix some $\lambda \in I$. $E_{t\wedge T}$ is exactly the process $(L_t)_{t\geq 0}$ defined in [4, Theorem VI-2] for $\mu_s = \exp(\lambda H_s)$ applied to the counting process $N_{t\wedge T}$ with compensator $\Lambda_{t\wedge T}$. As we verify all the assumptions of this theorem (with the notations of this theorem, μ_s is positive), we apply it and we obtain that $E_{t\wedge T}$ is a super-martingale. The result for stopping times is then obvious. ■

If N is a time Poisson process (i.e. Λ is deterministic), we can prove that $E_{t\wedge T}$ is a real martingale using Theorem II-8 of [4], with $H_s = f(s)$ deterministic. Therefore we recover the classical expression for the Laplace transform:

$$\forall t \geq 0, \mathbb{E}\left(e^{\lambda \int_0^t f(s) dN_s - d\Lambda_s}\right) = \exp\left(\int_0^t \phi(\lambda f(s)) d\Lambda_s\right).$$

In the other cases, the compensator is no longer deterministic, hence we cannot derive precise formula. This formula can be compared to its discrete time versions in [7, Lemma VII-2-8] or in [8]. These two versions can be applied to the discrete time martingale $\sum_{i=1}^n f(X_i)$ where the X_i's are i.i.d. and f with zero mean: we can then recover Bernstein's and Bennett's inequalities. We can do the same thing here.

Corollary 2. *For all positive λ such that $\int_0^T e^{\lambda H_s} d\Lambda_s$ is a.s. finite and for all positive ε*

$$\mathbb{P}\left(\sup_{[0,T]} Z_t \geq \varepsilon\right) \leq e^{-\lambda\varepsilon} \exp \left|\left| \int_0^T \phi(\lambda H_s) d\Lambda_s \right|\right|_\infty .$$

Proof. Let $\tau = \inf\{t \leq T / Z_t > \varepsilon\}$: it is a stopping time. We get by the positivity of ϕ that

$$\begin{aligned}
\mathbb{P}\left[\sup_{[0,T]} Z_t \geq \varepsilon\right] &= \mathbb{P}\left[Z_\tau \geq \varepsilon\right] \\
&= \mathbb{P}\left[\exp\left(\lambda Z_\tau - \int_0^\tau \phi(\lambda H_s) d\Lambda_s\right) \geq \exp\left(\lambda\varepsilon - \int_0^\tau \phi(\lambda H_s) d\Lambda_s\right)\right] \\
&\leq \mathbb{P}\left[\exp\left(\lambda Z_\tau - \int_0^\tau \phi(\lambda H_s) d\Lambda_s\right) \geq \exp\left(\lambda\varepsilon - \left|\left| \int_0^T \phi(\lambda H_s) d\Lambda_s \right|\right|_\infty\right)\right].
\end{aligned}$$

Then Markov inequality and Proposition 4 lead to the result. ∎

We only have to apply this result and optimizing it in λ to recover Theorem 1 and Proposition 1. In the both cases, the assumption "$\int_0^T e^{\lambda H_s} d\Lambda_s$ a.s. finite" is an obvious consequence of the assumptions of boundedness or the existence of moments for $(H_t)_{t\geq 0}$.

If the assumptions of boundedness or the existence of moments for $(H_t)_{t\geq 0}$ are verified just on an event $\mathcal{C}$, we can stop the martingale at a stopping time τ which is the infimum of the times t where one of the assumptions is wrong. Then we use the inequality for the stopped martingale. The result gives the exponential bound for the probability of $\{\sup Z \geq x\} \cap \mathcal{C} = \{\sup Z \geq x\} \cap \{\tau > T\}$ which is a subset of $\{\sup Z^\tau \geq x\}$.

3 Suprema

Now let $\{(H_{a,t})_{t\geq 0}, a \in \mathcal{A}\}$ be a countable family of locally bounded predictable processes. Let $(Z_t)_{t\geq 0}$ be

$$\forall t \geq 0, Z_t = \sup_{a\in\mathcal{A}} \left[\int_0^t H_{a,s} dM_s\right]. \tag{3.1}$$

Hence $(Z_t)_{t\geq 0}$ is an adapted process with bounded variations.

Under Assumption 1, the jumps of Z happen only when N jumps. Let T be a fixed positive number. For all t less than T, let us denote by $(T_i, 1 \leq i \leq n_t)$ the ordered jumps of N before t: there exists a.s. a finite number of these jumps as follows from Theorem II-8 (α) of [4] and Assumption 1. Consequently we can write:

$$\forall t \leq T, \quad Z_t = \sum_{T_i \leq t} [Z_{T_i} - Z_{T_i -}] + Z_{t-} - Z_{T_{n_t}} + \sum_{T_i \leq t} [Z_{T_i -} - Z_{T_{i-1}}] \text{ a.e.} \tag{3.2}$$

where $Z_{T_0} = Z_0 = 0$.

Lemma 1. *Assume $\mathcal{A} = \{1, ..., k\}$ finite. Let $i \geq 1$ be some integer. Let v be a real number in $]T_{i-1}; T_i[$. Then under Assumption 1, $Z_v - Z_{T_{i-1}} = -\int_{T_{i-1}}^{v} H_{\hat{a}_{s-},s} d\Lambda_s$, where $\hat{a}_{s-}$ is the first index where Z_{s-} is attained.*

Proof. Let us denote by $f(t)$ the Radon-Nikodym derivative $d\Lambda_t/dt$ which exists by Assumption 1. Then we can write for all v in $]T_{i-1}; T_i[$, $Z_v = \sup_{a\in\mathcal{A}} g_a(v)$ where $g_a(v) = \left[b_a + \int 1\!\!1_{]T_{i-1},v]} f_a(s)ds\right]$ and where $f_a(s) = -H_{a,s} f(s)$. As the g_a's are absolutely continuous and with normalized bounded variations and as $\mathcal{A}$ is finite, Z is also absolutely continuous and with normalized bounded variations. Consequently (see [12]), Z is almost surely differentiable and $Z_v = \int 1\!\!1_{]T_{i-1},v]} Z'_s ds$. So we have to compute its derivative to conclude. Let us restrict ourselves to the set of v's (with full measure) where $\forall a \in \mathcal{A}, g_a(v)' = f_a(v)$. For u also in $]T_{i-1}; T_i[$, we have the following inequalities:

$$\int_{]u,v]} f_{\hat{a}_u}(s)ds \leq Z_v - Z_u \leq \int_{]u,v]} f_{\hat{a}_v}(s)ds.$$

Let us divide by $v - u$. We do the proof for the left derivative. The same proof can be done for the right derivative. We take the limit when $u \uparrow v$ and we obtain:

$$f_{\hat{a}_{v-}}(v) \leq \liminf_{u\uparrow v} \frac{Z_v - Z_u}{v-u} \leq \limsup_{u\uparrow v} \frac{Z_v - Z_u}{v-u} \leq f_{\hat{a}_v}(v). \tag{3.3}$$

If the inequality is strict in (3.3) for v_0, it means that "$f_{\hat{a}_{v_0-}}(v_0) < f_{\hat{a}_{v_0}}(v_0)$". Hence we are in the following case: "$g_{\hat{a}_{v_0-}}(v_0) = g_{\hat{a}_{v_0}}(v_0)$ (else there is no reason to change the index where the supremum is attained) and in a neighborhood of the form $]v_0 - \varepsilon, v_0]$ for ε positive well chosen, we have "$g_{\hat{a}_{v_0-}}(u) > g_{\hat{a}_{v_0}}(u)$" because the functions are differentiable in v_0: v_0 is then isolated. (In order to be absolutely rigorous, we have to proceed by induction on the cardinality of $\mathcal{A}$.) Consequently, the set of v for which the inequality is strict has measure zero (it is countable). Then we get that almost everywhere "$Z'_v = f_{\hat{a}_{v-}}(v)$". This concludes the proof. ■

Using this lemma we get the following result.

Proposition 5. *Let T be a fixed positive number. Let $(Z_t)_{t\geq 0}$ be defined by (3.1). Under Assumptions 1, one has if $\mathcal{A}$ is finite:*

$$\forall 0 \leq t \leq T, \quad Z_t = \int_0^t \Delta Z(s) dN_s - \int_0^t H_{\hat{a}_{s-},s} d\Lambda_s \ a.s.$$

where $\Delta Z(s) = \sup_{a\in\mathcal{A}} \left[H_{a,s} + \int_0^{s-} H_{a,u} dM_u\right] - \sup_{a\in\mathcal{A}} \left[\int_0^{s-} H_{a,u} dM_u\right]$. Consequently, the compensator of $(Z_{t\wedge T})_{t\geq 0}$ is defined by

$$\forall t \geq 0, A_t = \int_0^{t\wedge T} \left[\Delta Z(s) - H_{\hat{a}_{s-},s}\right] d\Lambda_s.$$

If $\mathcal{A}$ is just countable, the compensator of $(Z_{t\wedge T})_{t\geq 0}$, $(A_t)_{t\geq 0}$ exists, is positive and nondecreasing and

$$\forall 0 \leq t \leq T, \quad Z_t - A_t = \int_0^t \Delta Z(s) dM_s.$$

Proof. Assume $\mathcal{A}$ finite. The first integral in Z_t is exactly the first part in (3.2). For the second part, all the differences are between two consecutive jumps and we can use the previous lemma. Moreover $\Delta Z(s)$ introduced in the proposition is predictable. The compensator is then obvious. As $\Delta Z - H_{\hat{a}_{s-},s}$ is positive and Λ nondecreasing, A is positive nondecreasing.

If $\mathcal{A}$ is just countable, $\mathcal{A}$ is an increasing union of finite sets B_n. Let us denote by Z^n the supremum over B_n instead of $\mathcal{A}$. As, for all n, B_n is finite, Z^n verifies the first part of the proposition. But, for all t less than T, $Z_t^n - \int_0^t \Delta Z^n(s)dN_s$ which is predictable, converges almost surely to $Z_t - \int_0^t \Delta Z(s)dN_s$ which is X_t: X_t is consequently also predictable. Then $A_t = \int_0^t \Delta Z(s)d\Lambda_s + X_t$ is the compensator of $Z_{t\wedge T}$ and it stays positive nondecreasing as a limit of positive nondecreasing functions. ■

To derive Proposition 3, it is sufficient to apply Proposition 1 or Corollary 1 to $Z - A$ noticing that ΔZ is bounded by $\sup_{a\in\mathcal{A}} |H_a|$.

4 Statistical applications

4.1 Statistical background

These exponential inequalities are useful to provide exponential deviations for the following χ^2-type statistics. Let T be a fixed positive real number and let $\{h_\lambda, \lambda \in m\}$ be a finite family of predictable processes. We set

$$\chi_T^2 = \sum_{\lambda\in m} \left(\int_0^T h_\lambda(t)dM_t \right)^2 . \tag{4.1}$$

This quantity appears naturally if we estimate the signal s by model selection in the white noise framework (see [3]). One has a model i.e. a finite dimensional linear subspace with orthonormal basis $\{\varphi_\lambda, \lambda \in m\}$ for the classical scalar product on $[0, T]$. The classical projection estimator on this subspace verifies that the square distance between the estimator and the true orthogonal projection of s, $\|s_m - \hat{s}_m\|^2$, is a χ_T^2 given by (4.1) with φ_λ instead of h_λ (i.e. deterministic functions) and dW the white noise instead of dM. In this case, this quantity is a real χ^2-statistics. The deviations of this quantity have to be controlled to prove adaptation properties for the model selection method. In the white noise framework, we can use the exponential inequalities available for χ^2-statistics.

If we estimate the density s from a n-sample by model selection, we can consider the same model as before. The distance $\|s_m - \hat{s}_m\|^2$ is also a χ^2-type statistics where $h_\lambda = \varphi_\lambda$, i.e. an orthonormal deterministic basis of the model for the classical scalar product. In this case, as in (1.3), dM is replaced by the centered empirical measure. In this context, L. Birgé and P. Massart use Talagrand's inequality to provide control on the χ^2-type statistics [2].

If we estimate the intensity s of a Poisson process N by model selection, we can consider always the same model as before. The distance $\|s_m - \hat{s}_m\|^2$ is always a χ^2-type statistics where dM is the centered process, and where $h_\lambda = \varphi_\lambda$, i.e. always an orthonormal deterministic family of $\mathbb{L}^2([0, 1], dt)$. In this case, we can use the concentration inequality of [9] (see Proposition 7) to control these distances, which give the same order of magnitude as Talagrand's inequality in the n-sample framework.

More generally, we can look at the Aalen multiplicative intensity model where the compensator of N verifies $d\Lambda = Ys(t)dt$, with Y predictable and known. For instance the censorship framework (see [1], for description of the framework) verifies this model. We can estimate the deterministic function s using the observations of the processes N and Y, by model selection (see [10]). In this case we are using a random scalar product $\int_0^t fgY\,dt$ instead of the classical one for the Poisson process when Y is constant. In this context, we cannot use exactly the same model as before but we can manage to get predictable h_λ's.

Indeed, if $\{\varphi_\lambda, \lambda \in m\}$ is an orthonormal family of $\mathbb{L}^2([0,1], dt)$ (typically histograms or Fourier basis), $\left\{h_\lambda = \varphi_\lambda/\sqrt{Y}, \lambda \in m\right\}$ becomes an orthonormal family for the random product (when Y is positive) and the h_λ's are predictable. The subspace generated by the h_λ's is used as model.

Our aim is now to provide a ready to use exponential inequality for χ_T.

4.2 A inequality which is ready for immediate application

To provide a concentration inequality for χ_T^2, we can remark that

$$\forall t \geq 0, \quad \chi_t = \sup_{\sum_{\lambda \in m} a_\lambda^2 = 1} \int_0^t \left(\sum_{\lambda \in m} a_\lambda h_\lambda(s) \right) dM_s. \tag{4.2}$$

Consequently we can use Proposition 3 on a countable dense subset of the unit ball of $\mathbb{R}^m$. But as we do not know in practice the compensator of χ_t we want to compare it to $\sqrt{C_t}$ where

$$\forall t \geq 0, \quad C_t = \sum_{\lambda \in m} \int_0^t h_\lambda(s)^2 d\Lambda_s, \tag{4.3}$$

is the compensator of χ_t^2. Finally we can obtained the forthcoming result.

Proposition 6. *Let T be a fixed positive real number. Let χ_T be defined by (4.1). Then, for all u positive, with probability larger than $1 - 2e^{-u}$,*

$$\chi_T - \sqrt{C_T} \leq 3\sqrt{2vu} + bu,$$

where

- *C_T is defined by (4.3);*
- *$v = \|C_T\|_\infty$;*
- *for all s less than T, $\sum_{\lambda \in m} h_\lambda^2(s) \leq b^2$.*

Proof. Let u be positive. First we can interpret χ_t as a supremum (see (4.2)). But moreover, we can take B a countable dense subset of the unit ball of $\mathbb{R}^m$ and say that

$$\chi_t = \sup_{a \in B} \int_0^t \left(\sum_{\lambda \in m} a_\lambda h_\lambda(s) \right) dM_s. \tag{4.4}$$

Then we can apply Proposition 3(a) with $H_a = \sum_{\lambda \in m} a_\lambda h_\lambda$. We obtain that $(\chi_{t\wedge T})_{t\geq 0}$ has a compensator $(A_t)_{t\geq 0}$ and

$$\mathbb{P}\left(\sup_{[0,T]} (\chi_t - A_t) \geq \sqrt{2vu} + \frac{b}{3}u \right) \leq e^{-u}.$$

We can replace the H_a by $-H_a$ to obtain

$$\mathbb{P}\left(\sup_{[0,T]}|\chi_t - A_t| \geq \sqrt{2vu} + \frac{b}{3}u\right) \leq 2e^{-u}.$$

Let $B_T = \sup_{[0,T]}|\chi_t - A_t|$. Now we must compare $(A_t)_{t\geq 0}$ and $(C_t)_{t\geq 0}$. One has for all t less than T:

$$\begin{aligned}
\chi_t^2 - A_t^2 &= (\chi_t - A_t)^2 + 2A_t(\chi_t - A_t) \\
&= (\chi_t - A_t)^2 + 2\int_0^t (\chi_{s-} - A_{s-})dA_s + 2\int_0^t A_s d(\chi_s - A_s).
\end{aligned}$$

But the first term has for compensator $\int_0^t (\Delta\chi)^2(s)d\Lambda_s$ and the last term is a martingale. Moreover $(A_t)_{t\geq 0}$ is predictable, thus we can take the compensator of the previous expression to obtain:

$$C_t - A_t^2 = \int_0^t (\Delta\chi)^2(s)d\Lambda_s + 2\int_0^t (\chi_{s-} - A_{s-})dA_s.$$

Consequently, one has:

$$\begin{aligned}
\chi_t - \sqrt{C_t} &= \chi_t - A_t + A_t - \sqrt{C_t} \\
&= \chi_t - A_t - \frac{C_t - A_t^2}{A_t + \sqrt{C_t}} \\
&= \chi_t - A_t - \frac{\int_0^t (\Delta\chi)^2(s)d\Lambda_s + 2\int_0^t (\chi_{s-} - A_{s-})dA_s}{A_t + \sqrt{C_t}}.
\end{aligned}$$

As $(A_t)_{t\geq 0}$ is positive and nondecreasing, one gets $\chi_t - C_t \leq 3B_T$, for all t less than T, which implies the result. ■

4.3 Orders of magnitude

Let us rewrite Corollary 1 of [9]:

Proposition 7. *Let N be a Poisson process on $(\mathbb{X}, \mathcal{X})$ with finite mean measure ν. Let $\{\psi_a, a \in A\}$ be a countable family of functions with values in $[-b, b]$. One considers*

$$Z = \sup_{a\in A}\left|\int_{\mathbb{X}} \psi_a(x)(dN_x - d\nu_x)\right| \text{ and } v_0 = \sup_{a\in A}\int_{\mathbb{X}} \psi_a^2(x)d\nu_x.$$

Then for any positive numbers u and ε:

$$P\big(Z \geq (1+\varepsilon)E(Z) + \sqrt{2\kappa v_0 u} + \kappa(\varepsilon)bu\big) \leq \exp(-u),$$

where $\kappa = 6$ and $\kappa(\varepsilon) = 1.25 + 32/\varepsilon$.

This inequality available for Poisson processes is exactly the same orders of magnitude as in Talagrand's inequality in the i.i.d. framework: the supremum lays outside the integral in v_0.

Let us compare it to Proposition 6 applied to the simplest Poisson process. We are in the case of s is a constant equal to 1, Y is a constant equal to n and T is 1 (i.e. $d\nu = nds$

is the mean measure of the Poisson process). This is the case for the aggregated process built from the sum of n i.i.d. homogeneous Poisson processes on $[0,1]$ with intensity 1.

Suppose the model is the set of histograms constructed on a regular partition m of $[0,1]$. Then the basis is deterministic and of the form $\sqrt{(D/n)}\mathbb{1}_I$ for I in m where D is the number of intervals in m.

If we apply Corollary 1 of [9] on $[0,1]$, to

$$\chi = \sup_{\sum_{I\in m} a_I^2=1} \int_0^1 \sum_{I\in m} a_I \sqrt{\frac{D}{n}}\mathbb{1}_I(dN_s - nds) = \sqrt{\sum_{I\in m} \frac{D}{n}\left(N_I - \frac{n}{D}\right)^2},$$

we obtain, for all u and ε positive numbers,

$$\mathbb{P}\left(\chi \geq (1+\varepsilon)\sqrt{D} + \sqrt{2\kappa u} + \kappa(\varepsilon)\sqrt{\frac{D}{n}u}\right) \leq \exp -u. \tag{4.5}$$

But, if we apply Proposition 6, we obtain, for all u positive:

$$\mathbb{P}\left(\chi \geq \sqrt{D} + 3\sqrt{2Du} + \sqrt{\frac{D}{n}u}\right) \leq 2\exp -u. \tag{4.6}$$

The variance term (the factor of $\sqrt{u}$) in (4.6) is bigger than the corresponding term in (4.5). It is of the same order than the expectation ($\sqrt{D}$). We must mention that D can become very big: it can be as large as n for proper model and to have good estimation n grows to infinity. Consequently, (4.5) gives an order of magnitude of $\sqrt{D}$ when (4.6) gives an order of magnitude of $\sqrt{Du}$. In this sense, Inequality (4.5) is better than (4.6).

But for more general Aalen multiplicative intensity processes, Y is no longer constant: it often decreases and can become very small. When t tends to 1, Y_t is equal to 1 in the right censorship framework for instance but at 0, Y_0 is n, the number of observations. In this case, the third linear term becomes of order $u\sqrt{D}$. Consequently, even if we were able to improve the behavior of the quadratic term in the general case, we would not change the order of magnitude given by these type of inequality for Aalen multiplicative intensity processes.

Conclusion

We have proved an inequality which is the equivalent of Talagrand's inequality in the n-sample framework even if we increase the order of magnitude of the quadratic term. This is especially useful to prove oracle inequalities in model selection when we are dealing with counting processes which are more intricate than Poisson processes as are the processes with Aalen multiplicative intensity.

References

[1] P.K. Andersen, O. Borgan, R. Gill, and N. Keiding. *Statistical Models Based on Counting Processes.* Springer Series in Statistics, 1993.

[2] L. Birgé and P. Massart. From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pages 55–87. Springer, New York, 1997.

[3] L. Birgé and P. Massart. Gaussian model selection. *Journal of the European Mathematical Society*, 2001.

[4] P. Brémaud. *Point processes and queues.* Springer-Verlag, 1981.

[5] C. Houdré and N. Privault. Concentration and deviation inequalities in infinite dimensions via covariance representations. *To appear in Bernoulli*, 2002.

[6] O. Kallenberg. *Foundations of modern probability.* Springer, 1997.

[7] J. Neveu. *Discrete-Parameter Martingales. Translated by T. P. Speed.* American Elsevier Publishing Company, Inc., 1975.

[8] I. Pinelis. Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Prob.*, 22:1679–1706, 1994.

[9] P. Reynaud-Bouret. Concentration inequalities for inhomogeneous Poisson processes and adaptive estimation of the intensity. Technical Report 18, Université de Paris-Sud, 2001.

[10] P. Reynaud-Bouret. *Estimation adaptative de l'intensité de certains processus ponctuels par sélection de modèle.* PhD thesis, Université Paris XI, UFR Orsay, 2002.

[11] E. Rio. Une inégalité de Bennett pour les maxima de processus empiriques. Technical report, Université de Versailles-St Quentin en Yvelynes, 2001.

[12] W. Rudin. *Analyse réelle et complexe.* Masson, 1987.

[13] P.-M. Samson. Concentration of measure inequalities for Markov chains and ϕ-mixing processes. *Ann. Prob.*, 2000.

[14] Sara van de Geer. Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes. *Ann. Stat.*, 23(5):1779–1801, 1995.

[15] L. Wu. A new modified logarithmic Sobolev inequality for Poisson point process and several applications. *Probability Theory and Related Fields*, 2000.